

SUSMAN GODFREY L.L.P.

January 16, 2025

Hon. Ona T. Wang  
Southern District of New York  
500 Pearl Street  
New York, New York 10007

Re: *The New York Times Company v. Microsoft Corporation, et al.*,  
No.: 23-cv-11195-SHS: Opp. to OpenAI's Motion to Compel, ECF [393](#)

Dear Magistrate Judge Wang:

At the December 3 conference, the Court made clear it expected the parties to meet and confer further regarding OpenAI's request to add David Rubin, Michael Greenspon, Hannah Masuga, and Sam Dolnick as custodians. Rather than doing so, and without any attempt to compromise or narrow the issues in dispute, OpenAI notified Plaintiffs on January 9, that it planned to "renew" its motion regarding these four individuals. It further seeks documents from two additional custodians (Justin Heideman and Elijah Soria) that are duplicative of existing custodians, and to expand the date range for all Times custodians by four years, all without regard to relevance, burden, or proportionality. The four custodians OpenAI previously requested are no more appropriate now than they were in December; and OpenAI's rationale for requesting these further custodians includes to obtain discovery into "The Times's AI strategy" and ways it can use AI, topics the Court has already ruled to be irrelevant.

OpenAI's motion discusses at length The Times's market cap, the size of its IT department, and the considerable resources it has spent on this litigation, but says little about why the discovery it seeks is proportionate and non-duplicative of existing custodians. The Court should deny OpenAI's motion at ECF [393](#).

**1. OpenAI Provides No Basis for Compelling Previously-Sought Custodians.**

OpenAI misrepresents The Times's arguments against adding Rubin, Greenspon, Masuga, and Dolnick as custodians, claiming that at the December 3 hearing The Times's "principle argument . . . was burden." ECF [393](#) at 1. That is incorrect. The Times argued that OpenAI has not demonstrated why the addition of these custodians is proportionate to the needs of the case given that The Times has designated numerous custodians with documents relevant to the discovery OpenAI claims to need. *See* ECF 330; ECF 393-5 (12/3 Hearing Trans.) at 37:1–21. Its renewed motion and the attached exhibits still fail to do so.

For example, OpenAI describes Dolnick and Rubin as members of "a core group tasked with prepping for meetings with OpenAI and setting The Times's AI strategy," ECF [393](#) at 1 n.2 (citing Ex. D); but, as The Times previously explained, all relevant licensing work related to generative AI runs through existing custodians Sam Felix and Rebecca Grossman-Cohen, who led The Times's negotiations with OpenAI and Microsoft, ECF 330 at 2. Indeed, four members of the "core group" listed on this email, as OpenAI frames them, are already custodians (Poferl, Hardiman, Grossman-Cohen, and Felix). OpenAI makes no attempt to explain why these custodians are insufficient. And OpenAI's argument that it is entitled to discovery into The

January 15, 2025  
Page 2

Times's broader "AI strategy" has already been rejected in this Court's November 22 and December 2 Orders. ECFs 344, 353, 355.

Similarly, OpenAI attaches to its motion two meeting invitations, though it makes no attempt to explain how they support its argument. One of those includes four current custodians (Grossman-Cohen, Sobel, Hardiman, and Pofert (ECF [393](#), Ex. B)), and the other includes five (Levien, Grossman-Cohen, Sulzberger, Hardiman, and Brayton (ECF [393](#), Ex. C)). If OpenAI's point is that it needs custodians who participated in these calls, it has them in spades.

Finally, OpenAI cites to an email from Greenspon in which he states that an upcoming meeting will consider "how licensing can use AI to drive our business" as support for adding Greenspon as a custodian. ECF [393](#) at 1 n.2 (citing Ex. A). This argument again disregards the Court's ruling on the scope of permissible discovery. ECF 344, at 5.

OpenAI cites no new reason to add Masuga, and The Times has already explained why doing so is not proportionate. *See* ECF 330 at 2–3.

Without a substantive basis for requesting the addition of these custodians, OpenAI resorts to mischaracterizing The Times's positions and discovery efforts and taking statements from The Times's CTO Jason Sobel out of context. This is an unnecessary side show. The Times explained in December that collecting from these custodians would delay its collection of other agreed-upon custodians. The Times has completed those custodial collections for most sources<sup>1</sup>; the parties have been actively negotiating search terms and custodians; and The Times is working through the more than 100,000 custodial documents currently batched for review—a number that will continue to expand once the parties reach agreement on the additional search terms currently under negotiation. Nothing about The Times's conduct (including the millions of dollars, highlighted by OpenAI, that The Times has spent to litigate this case) supports OpenAI's accusations of "shirking." It also does not support compelling the addition of four duplicative and unnecessary custodians. The Court should deny OpenAI's request that it do so.

## **2. OpenAI Does Not Provide a Basis to Add Heideman and Soria.**

OpenAI's request to add two software engineers, Heideman and Soria, as custodians for search terms targeting the "operation of The Times's paywall, webmaster controls, and exclusion protocols" should be denied based on the existing custodians and the data to which OpenAI has access. In particular, existing custodians Hannah Pofert (Chief Data Officer & Head of Audience), Christine Liang (Senior Director of Search Engine Optimization), and Sam Felix (VP & Head of Platform Strategy and Strategic Generative AI Partnerships) are knowledgeable on the same topics (and were disclosed by The Times as such, *see* 393-9 at Resp. 7 & 8).

OpenAI highlights its claim that "The Times has not yet produced 'documents sufficient to show the time periods during which The Times implemented Exclusion Protocols on [its] website,'" arguing that this "issue can be traced to The Times's designation of custodians." ECF [393](#). OpenAI is mistaken for multiple reasons. As an initial matter, The Times is continuing to

---

<sup>1</sup> The Times ran into technical issues with its Slack pulls that it has worked with its outside vendor to address.

January 15, 2025  
Page 3

produce documents from the agreed-upon custodians. But more importantly, it is grossly inefficient to try to discover the implementation of The Times’s robots.txt protocol via the “day-to-day discussions” OpenAI seeks in its motion. The Times has already agreed to produce non-custodial documents reflecting this information. For instance, The Times’s products and engineering teams use a software program, Jira, to track and assign work, with each task being represented by a “ticket.” The Times has produced the Jira tickets hitting on “robots.txt” going back to 2012, which speak to things like adding new robots files to subdomains, adding new disallow/allow rules, and disallowing specific crawlers. Ex. A (excerpt of NYT\_00692315). And finally, The Times’s historic robots.txt pages are publicly available: searching the Internet Archive for “nytimes.com/robots.txt” pulls up tens of thousands of screenshots of The Times’s robots.txt page, spanning from January 16, 1999 to the present. *See, e.g.*, Ex. B (1/16/99 screenshot).

### **3. A Blanket Collection from 2018 is Disproportionate to the Needs of the Case.**

The parties first negotiated search terms, custodians, and date ranges for collection in July 2024. As part of those negotiations, The Times represented that it would collect documents from its custodians beginning in 2022 (ChatGPT was publicly released in November 2022), and Defendants did not object to that proposal. The Times then collected documents from its custodians from the start of 2022 through May 1, 2024. Subsequently, months later, OpenAI sought to reopen negotiations regarding the timeframe for collection. It demanded that The Times go back and re-collect documents from specific existing custodians, and The Times did so. OpenAI’s yet further demand that The Times conduct a blanket collection of an additional four years of its custodians’ documents is unjustified.

OpenAI claims it is entitled to “communications” regarding “OpenAI’s alleged misconduct” prior to November 2022 from all “relevant custodians,” without specifying who it believes those are. As OpenAI concedes, The Times has agreed to produce documents going back to 2018 for several of its high-ranking employees. ECF [393](#) at 3 n.7. To avoid the significant burden of re-collecting from existing custodians, The Times further offered to work with other requested custodians to directly run searches in their email for terms like “OpenAI” and “ChatGPT” for the period prior to 2022. OpenAI did not respond to that offer. The Times maintains that a full custodial collection for all or most custodians extending four years prior to the public release of ChatGPT—which made OpenAI a “viral sensation,” Compl. ¶61—makes no sense.

With respect to evidence of impact on The Times, *OpenAI* has documents relevant to its misconduct from well before 2022, when it surreptitiously used millions of The Times’s works to develop its products; but, because much of this unlawful activity came before the public release of OpenAI’s GPT product in 2022, The Times is not likely to have custodial documents relevant to impact from this earlier period. The Times is already producing non-custodial documents related to the harm it incurred—documents related to revenue, subscriptions, and web traffic—going back to 2018 and, in some instances 2015. With regard to custodians’ communications, the Court has already denied Defendants’ motion to compel additional communications on exactly these topics. *See* ECFs 351–352 (denying motions to compel communications related to, e.g., changes to The Times’ revenue, subscriptions, and web traffic).

January 15, 2025  
Page 4

Respectfully submitted,

/s/ Ian B. Crosby  
Ian B. Crosby  
Susman Godfrey L.L.P.

/s/ Steven Lieberman  
Steven Lieberman  
Rothwell, Figg, Ernst & Manbeck

cc: All Counsel of Record (via ECF)